New Eye-Tracking Techniques May Revolutionize Mental Health Screening

Laurent Itti[1,*]
[1]Computer Science Department, Psychology Department, and Neuroscience Graduate Program, University of Southern California, 3641 Watt Way, HNB-07A, Los Angeles, CA 90089-2520, USA
*Correspondence: itti@usc.edu
http://dx.doi.org/10.1016/j.neuron.2015.10.033

Visually-guided behavior recruits a network of brain regions so extensive that it is often affected by neuropsychiatric disorders, producing measurable atypical oculomotor signatures. Wang et al. (2015) combine eye tracking with computational attention models to decipher the neurobehavioral signature of autism.

Imagine that you are waiting for a prescription to be filled at your local pharmacy. Today, many pharmacies in the United States of America provide free blood pressure monitors that you can use as a rapid health indicator while you wait. You simply place your arm in the pressure band, press a button, and observe the readings. If the readings are high, it may be a good idea to check with your doctor whether any corrective action should be taken. What if similar devices could be made available for the evaluation of mental health?

Recent progress in eye-tracking techniques is opening new avenues for quantitative, objective, simple, inexpensive, and rapid evaluation of mental health, as shown in this issue by the study of Wang et al. (2015). The starting premise is that the visual attention and eye movement networks are so pervasive in the human brain (Corbetta and Shulman, 2002; Miller and Buschman, 2013) that many neurodevelopmental and neurodegenerative disorders may affect their functioning, resulting in quantifiable alterations of eye movement behavior (Leigh and Zee, 2015). Indeed, the control of attention and gaze involves not only occipital (early vision), temporal (high-level object vision), parietal (spatial vision and attention), and frontal (goal-driven vision) cortices, but also the limbic system, reward systems, and deep-brain nuclei including the thalamus and the superior colliculus (Baluch and Itti, 2011; Gottlieb et al., 2014) (Figure 1A). Consequently, many previous studies have demonstrated differences in saccadic reaction time, in saccade and fixation metrics, and in error patterns during visually-guided behavior, for a wide range of neurobehavioral disorders. Most studies to date have used structured laboratory tasks and stimuli. A notable example is the anti-saccade task, where a peripheral target suddenly appears on a blank screen and the task is to refrain from looking at it, and to instead look in the opposite direction where there is nothing on the display (Munoz and Everling, 2004). Patients show either markedly slower reaction times, increased error rates, or both, compared to controls, as tested with attention deficit hyperactivity disorder (ADHD), Tourette's syndrome, Parkinson's disease, and schizophrenia (Munoz and Everling, 2004). The same eye movement tasks can be used to monitor development and maturation, even in control subjects (Luna et al., 2008). This general approach already provides a valuable complement to more conventional neuropsychiatric assessment using questionnaires and clinical evaluations, especially for those disorders for which a clear chemical, genetic, morphological, physiological, or histological biomarker has not yet been identified.

The study of Wang et al. (2015) focuses on autism spectrum disorder (ASD) and replaces structured tasks and laboratory stimuli—previously extensively studied in ASD (e.g., Takarae et al., 2004)—by simple free viewing of natural images. This presents a number of advantages that have recently been noted in the literature (Tseng et al., 2013): the technique is applicable to very young children or to any individual (or animal) who may not understand or be interested in complying with any task instruction, the stimuli present a wider range of visual attributes (including low-level features, such as many different textures, colors, and shapes, but also, of particular interest in Wang et al.'s study, many different kinds of objects and many different semantic valences of items or actors in the scenes), and the amount of information collected per unit time is higher than for typical laboratory tasks based on a trial-by-trial structure. But this comes at the cost of significantly complicating the ensuing data analysis. Indeed, the inter-observer variability in gaze patterns during free viewing of natural scenes is large even within control populations (Mannan et al., 1995; also note the spread of fixations for some stimuli in Figure 1 of Wang et al., 2015), the stimuli are highly complex and cannot at present be fully characterized or codified by existing theories of object perception or visual scene understanding; the open-ended nature of the free-viewing task invites further variability due to cultural, gender, and other individual differences; and eye movements are recorded on a continuous basis as opposed to well-defined, discrete trial-by-trial episodes.

As a result, while obvious differences in fixation preferences may be notable between patient and control groups on some images (Figure 1 of Wang et al., 2015), quantifying those differences in terms of possible differences in attention allocation toward different attributes of the stimuli is not trivial. This is where new computational models and machine learning tools can help. Here, Wang et al. (2015) develop an elegant three-stage visual saliency model, which produces a topographic activation map for each natural image in their stimulus set. The map highlights locations in the image that are more conspicuous and hence more likely to be looked at, at least by

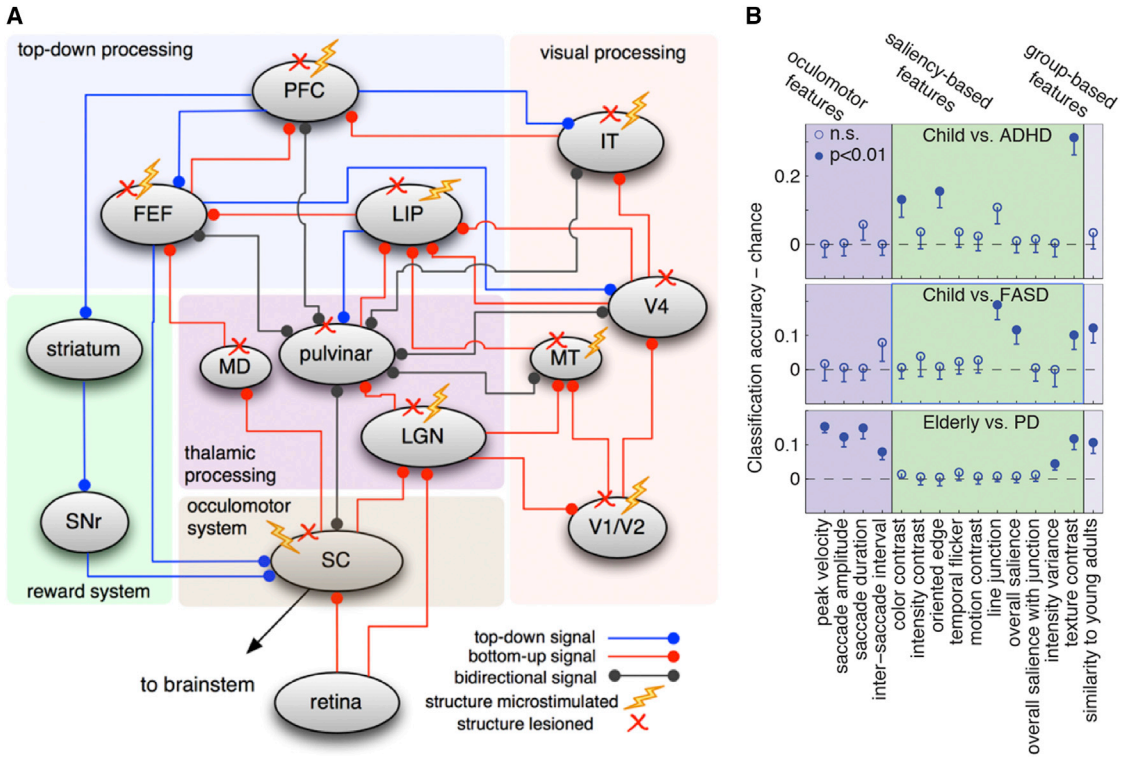

Figure 1. Brain Circuit of Attention and Behavioral Signatures of Disorders

(A) Signals and brain structures that have been implicated in attention and gaze control. The flash symbol indicates that a structure has been microstimulated and an X indicates that it has been lesioned in previous studies, to characterize its role in the circuit. The connections show the most likely type of signal being transmitted between two structures; top-down ([TD]; goal driven) signals in blue, bottom-up ([BU]; stimulus driven) signals in red, and bidirectional signals in gray. Abbreviations: SC, superior colliculus; SNr, substantia nigra pars reticulata; MD, mediodorsal thalamus; LGN, lateral geniculate nucleus; IT, inferotemporal cortex; MT, middle temporal area; LIP, lateral intraparietal area; FEF, frontal eye fields; PFC, prefrontal cortex. Reproduced with permission from Baluch and Itti (2011).

(B) Signatures of three different disorders—ADHD and FASD in children, as well as Parkinson's disease (PD) in elderly—obtained through eye tracking while participants freely watched natural video clips. The signatures show strikingly distinct, quantitative, and objective patterns of atypical deployment of gaze, here along 15 dimensions from three broad categories: oculomotor (saccade and fixation metrics), saliency based (attention to visual features of the stimuli), and group-based (correlation with gaze patterns of control young adults). Such signature or behavioral biomarker can be computed for any new individual and then classified using machine learning systems into the most likely patient or control group. Error bars indicate 95% confidence intervals after Bonferroni correction. Significance level: $p < 0.01$, one-tailed paired t-test. Adapted from Tseng et al. (2013).

control observers. Saliency maps are typically constructed as the weighted sum or aggregation of several feature maps, each sensitive to a particular visual attribute (e.g., color contrast or oriented edges; Itti et al., 1998). Using a standard machine learning system (support vector machine [SVM]), Wang et al. (2015) learn the relative weights of different features that contribute to saliency so as to maximize the agreement between model saliency maps and recorded human fixations, separately for the control and patient groups. Differences in the learned weights between patient and control groups indicate different levels of preference or different degrees of attractiveness of the features across the two groups. Wang et al., 2015 (see Figure 2) indeed report a striking set of differences between ASD patients and controls: model weights for ASD patients were higher for pixel-level salience, for the background of the scene, and for the image center, but lower for objects and items with semantic valence, such as faces or items being looked at by persons or animals in the scene. Repeating the analysis for individual saccades as they developed over the time spent scanning an image revealed, for both groups, a general decrease in the weights of low-level features and an increase in weights of object and semantic features (Figure 3 of Wang et al., 2015), suggesting a progressively lower influence of image-based or bottom-up features and higher influence of top-down factors in guiding gaze as time develops (as already noted previously with simpler models; e.g., Parkhurst et al., 2002). An interesting new result is the relatively lower decrease in weights for low-level features, and relatively lower increase in weights for object-level and semantic features, for ASD patients compared to controls (Figure 3 of Wang et al., 2015). These differences in model weights were corroborated in a model-free analysis that demonstrated fewer and slower saccades toward semantic objects for the ASD group, more fixations in the background, and longer fixations over background and other objects (Figures 4D–4F of Wang et al., 2015). Among the semantic features, motion, smell, and touch features had lower model weights in ASD patients compared to controls, as well as faces, for later fixations in the scanpaths (Figures 5 and S5E of Wang et al., 2015). Finally, although

more salient faces and text elements attracted more fixations overall, there was no difference between ASD and controls in this dimension (Figure 6 of Wang et al., 2015). Overall, the study reveals a complex pattern of differences between the ASD and control groups, with detailed ramifications into the nature of the items gazed at and the time at which they were gazed at. Because it is complex, this pattern may be viewed as lacking a simple interpretation and may be difficult to directly link to phenotype and underlying neurophysiology. For example, it is not true that ASD patients strongly avoided human faces and locations gazed at by humans or animals in the images, since the weights for these features are non-zero and, overall, not significantly lower for ASD patients than for controls (Figure 5 of Wang et al., 2015). This makes it difficult to directly translate the findings into neurological terms or into an interpretation of what functional brain mechanism differences may exist between the patient and control groups.

Yet, just because an elevated blood pressure reading does not fully explain what is abnormal in the circulatory system does not render it useless, and likewise with eye-tracking measures such as those of Wang et al., 2015. Although complex patterns of differences may not directly pinpoint which brain areas or functions are affected by a disease, they can be used as characteristic behavioral biomarkers—or behavioral biometric signatures—of particular disorders. Such signatures can support screening and differentiations among patients, not only at the level of group-based statistical effects (as shown by Wang et al., 2015), but also possibly for individual persons. For example, using a similar approach but with video rather than static stimuli, Tseng et al. (2013) were able to build a three-way classifier that could differentiate between children with ADHD, fetal alcohol spectrum disorder ([FASD], sometimes comorbid with ADHD), and controls, well above chance, from eye movements recorded over 15 min of watching television (Figure 1B). Applying the same methods but at the other end of the age spectrum, the classifier was 90% correct at differentiating Parkinson's disease patients from elderly controls. Thus, a bright future seems to lie ahead for approaches like those described in this issue by Wang et al., 2015. A key issue for the immediate future is that these techniques should be shown to be equally or more sensitive and specific than existing approaches, for example through longitudinal studies (Magiati et al., 2014) of large cohorts of initially undiagnosed individuals, some of whom may later develop a disorder that had been predicted by the model-based analysis. If that is the case, maybe some day in the not so distant future simple mental health assessment machines based on eye tracking may come to existence, possibly at your nearby pharmacy (Tseng et al., 2014).

ACKNOWLEDGMENTS

Professor Itti's research is supported by the National Science Foundation (grant numbers CCF-1317433 and CNS-1545089), the Army Research Office (W911NF-12-1-0433), and the Office of Naval Research (N00014-13-1-0563). The author affirms that the views expressed herein are solely his own and do not represent the views of the United States government or any agency thereof.

REFERENCES

Baluch, F., and Itti, L. (2011). Trends Neurosci. *34*, 210–224.

Corbetta, M., and Shulman, G.L. (2002). Nat. Rev. Neurosci. *3*, 201–215.

Gottlieb, J., Hayhoe, M., Hikosaka, O., and Rangel, A. (2014). J. Neurosci. *34*, 15497–15504.

Itti, L., Koch, C., and Niebur, E. (1998). IEEE Trans. Pattern Anal. Mach. Intell. *20*, 1254–1259.

Leigh, J.R., and Zee, D.S. (2015). The Neurology of Eye Movements, Fifth Edition (Oxford University Press).

Luna, B., Velanova, K., and Geier, C.F. (2008). Brain Cogn. *68*, 293–308.

Magiati, I., Tay, X.W., and Howlin, P. (2014). Clin. Psychol. Rev. *34*, 73–86.

Mannan, S., Ruddock, K.H., and Wooding, D.S. (1995). Spat. Vis. *9*, 363–386.

Miller, E.K., and Buschman, T.J. (2013). Curr. Opin. Neurobiol. *23*, 216–222.

Munoz, D.P., and Everling, S. (2004). Nat. Rev. Neurosci. *5*, 218–228.

Parkhurst, D., Law, K., and Niebur, E. (2002). Vision Res. *42*, 107–123.

Takarae, Y., Minshew, N.J., Luna, B., Krisky, C.M., and Sweeney, J.A. (2004). Brain *127*, 2584–2594.

Tseng, P.-H., Cameron, I.G.M., Pari, G., Reynolds, J.N., Munoz, D.P., and Itti, L. (2013). J. Neurol. *260*, 275–284.

Tseng, P.-H., Cameron, I.G.M., Munoz, D.P., and Itti, L. (2014). United States patent number 8,808,195 B2, filed Jan 15, 2010 following provisional application 61/145,011, filed Jan 15, 2009. Issued Aug 19, 2014.

Wang, S., Jiang, M., Duchesne, X.M., Laugeson, E.A., Kennedy, D.P., Adolphs, R., and Zhao, Q. (2015). Neuron *88*, this issue, 604–616.